Exam #1

Economics 435: Quantitative Methods

Spring 2006

1 Starter questions (20 points)

a) Let x and y be random variables, and let a, b, c, d be constants. Find $cov(ax, b + cx + dy)$.

b) Suppose that we have estimated a standard MLR model by ordinary least squares, and the 95% confidence interval for β_1 is $[0.23, 0.77]$.

- Would we reject the null hypothesis that $\beta_1 = 0$ at a 5% significance level?
- Would we reject the null hypothesis that $\beta_1 = 0$ at a 10% significance level?
- What is the actual value of $\hat{\beta}_1$?

c) Prove that $E(u|x) = 0$ implies that $E(u) = 0$.

d) Prove that $E(u|x) = 0$ implies that $cov(u, x) = 0$.

2 Biased sample selection (40 points)

Suppose that:

$$y = \beta_0 + \beta_1 x + u$$
$$E(u|x) = 0$$

You have a data set with observations on (y, x) but it is not a true random[1] sample.

They way we will model this is by imagining that there is an underlying random sample, but each individual in the sample is characterized by a binary random variable s such that we only observe (x_i, y_i) if $s_i = 1$. We will suppose that we know the selection mechanism itself, i.e.:

$$\Pr(s = 1 | x \in A \cap y \in B) \text{ for all possible } A, B$$

We will also suppose (to make the math easier) that x and y are discrete random variables.

In this problem we will establish conditions under which we can consistently estimate β_1.

a) First, note that if we can consistently estimate

$$\begin{aligned} Y_0 &\equiv E(y|x = X_0) \\ Y_1 &\equiv E(y|x = X_1) \end{aligned}$$

[1] One example where this might occur is oversampling, in which researchers intentionally sample a disproportionate number of respondents meeting certain characteristics (e.g. rural households, households in small states or provinces, low-income households, etc.)

for any two distinct x-values $X_0 \neq X_1$, we can consistently estimate β_1. Let $\hat{Y}_0$ and $\hat{Y}_1$ be consistent estimators of Y_0 and Y_1. Find a consistent estimator of β_1 as a function of $(X_0, X_1, \hat{Y}_0, \hat{Y}_1)$.

b) The analog principle implies that we can consistently estimate $E(y|x = X_0)$ if we can consistently estimate $\Pr(y = Y|x = X_0)$ for all possible values of Y. The standard estimate for a random sample would be to take all observations in the data in which $x_i = X_0$, then calculate the proportion of those observations in which $y_i = Y$ holds. In our setting this would be $\hat{P}_Y$, where:

$$\hat{P}_Y \equiv \frac{\frac{1}{n}\sum_{i=1}^n I(y_i = Y \cap x_i = X_0 \cap s_i = 1)}{\frac{1}{n}\sum_{i=1}^n I(x_i = X_0 \cap s_i = 1)}$$

Show that:

$$\text{plim } \hat{P}_Y = \Pr(y = Y|x = X_0, s = 1)$$

c) Unfortunately it is not necessarily the case that $\Pr(y = Y|x = X_0) = \Pr(y = Y|x = X_0, s = 1)$, so $\hat{P}_Y$ is not necessarily a consistent estimator of $\Pr(y = Y|x = X_0)$.

Use Bayes' Law[2] to show that:

$$\Pr(y = Y|x = X_0) = \frac{\Pr(y = Y|x = X_0, s = 1)\Pr(s = 1|x = X_0)}{\Pr(s = 1|x = X_0, y = Y)} \tag{1}$$

d) Using equation (1), determine a condition on the selection rule under which:

$$\text{plim } \hat{P}_Y = \Pr(y = Y|x = X_0)$$

Under this condition, known as *exogenous selection* or *ignorable selection* we can ignore the nonrandom selection issue and consistently estimate β_1 by OLS.

e) Using equation (1), determine a condition on the selection rule under which $\hat{P}_Y$ can be used (after an adjustment for the selection rule – remember that we know the selection rule) to construct a consistent estimate of $\Pr(y = Y|x = X_0)$.

Under this condition, it is possible to construct a consistent estimate of β_1, though it is not necessarily the OLS estimator.

3 Measurement error (40 points)

Often variables are measured imprecisely. This question develops the implications of measurement error for the properties of our regression coefficients. Suppose that:

$$\begin{aligned} y &= \beta_0 + \beta_1 x + u \\ E(u|x) &= 0 \end{aligned}$$

and that you have a random sample of data on $(\tilde{y}, \tilde{x})$ where

$$\begin{aligned} \tilde{y} &= y + e_y \\ \tilde{x} &= x + e_x \end{aligned}$$

[2] If you didn't write down Bayes' Law, or if you're not sure how to deal with the "$x = X_0$" part, here's a version of the law you will find useful:

$$\Pr(A|B \cap C) = \frac{\Pr(B|A \cap C)\Pr(A|C)}{\Pr(B|C)}$$

where A, B, and C are any events.

the unobserved variables e_y and e_x represent measurement error in y and x respectively. We will assume that the measurement error satisfies:

$$E(e_x|x, y, u) = E(e_y|x, y, u) = 0$$

That is, the measurement error is mean-independent of the true values. Note that this assumption implies:

$$cov(e_x, x) = cov(e_x, u) = cov(e_y, x) = cov(e_y, u) = 0$$

We will also assume that x has nonzero sample variance in our data.

Let $\hat{\beta}_1$ be the coefficient from an OLS regression of $\tilde{y}$ on $\tilde{x}$, i.e.:

$$\hat{\beta}_1 \equiv \frac{\hat{cov}(\tilde{x}, \tilde{y})}{\hat{var}(\tilde{x})}$$

Let b_1 be the probability limit of $\hat{\beta}_1$

a) Find b_1 as a function of β_1, $var(x)$, $var(e_x)$, $var(e_y)$, and $cov(e_x, e_y)$. You will use this result to answer the rest of the question, so be careful!

b) Suppose that x is measured with error but y is not, i.e.

$$\begin{aligned} var(e_x) &> 0 \\ var(e_y) &= 0 \end{aligned}$$

Note that $var(e_y) = 0$ also implies $cov(e_x, e_y) = 0$.

- Does $b_1 = \beta_1$?
- Do b_1 and β_1 have the same sign?
- Suppose that β_1 is positive. Is $b_1 < \beta_1$?

Possible answers are "Yes", "No", and "Uncertain".

c) Suppose that y is measured with error but x is not, i.e.

$$\begin{aligned} var(e_x) &= 0 \\ var(e_y) &> 0 \end{aligned}$$

Note that $var(e_x) = 0$ also implies $cov(e_x, e_y) = 0$.

- Does $b_1 = \beta_1$?
- Do b_1 and β_1 have the same sign?
- Suppose that β_1 is positive. Is $b_1 < \beta_1$?

Possible answers are "Yes", "No", and "Uncertain".

d) Suppose that x and y are both measured with error, but the measurement error is uncorrelated across the two variables, i.e.

$$\begin{aligned} var(e_x) &> 0 \\ var(e_y) &> 0 \\ cov(e_x, e_y) &= 0 \end{aligned}$$

- Does $b_1 = \beta_1$?
- Do b_1 and β_1 have the same sign?
- Suppose that β_1 is positive. Is $b_1 < \beta_1$?

Possible answers are "Yes", "No", and "Uncertain".

e) Suppose that x and y are both measured with error, and the measurement error is positively correlated across the two variables, i.e.

$$\begin{aligned} var(e_x) &> 0 \\ var(e_y) &> 0 \\ cov(e_x, e_y) &> 0 \end{aligned}$$

- Does $b_1 = \beta_1$?
- Do b_1 and β_1 have the same sign?
- Suppose that β_1 is positive. Is $b_1 < \beta_1$?

Possible answers are "Yes", "No", and "Uncertain".

f) Suppose that x and y are both measured with error, and the measurement error is negatively correlated across the two variables, i.e.

$$\begin{aligned} var(e_x) &> 0 \\ var(e_y) &> 0 \\ cov(e_x, e_y) &< 0 \end{aligned}$$

- Does $b_1 = \beta_1$?
- Do b_1 and β_1 have the same sign?
- Suppose that β_1 is positive. Is $b_1 < \beta_1$?

Possible answers are "Yes", "No", and "Uncertain".